LA-UR-04-1009

Approved for public release; distribution is unlimited.

Title: THE ROLE OF RNA EDITING IN DYNAMIC ENVIRONMENTS

Author(s): Luis M. Rocha, 121983, CCS-3
Chien-Feng Huang,185160, CCS-3

Submitted to: Ninth International Conference on the Simulation and Synthesis of Living Systems

The Role of RNA Editing in Dynamic Environments

Luis M. Rocha and Chien-Feng Huang

Modeling, Algorithms and Informatics Group (CCS-3)
Computer and Computational Sciences Division
Los Alamos National Laboratory, MS B256
Los Alamos, NM 87545, USA
rocha@lanl.gov and cfhuang@lanl.gov

Abstract

This paper presents a computational methodology based on Genetic Algorithms with Genotype Editing (GAE) for investigating the role of RNA editing in dynamic environments. This model is constructed based on several genetic editing characteristics that are gleaned from the RNA editing system as observed in several organisms. We have previously expanded the traditional Genetic Algorithm (GA) with artificial editing mechanisms (Rocha, 1995, 1997), and studied the benefits of including straightforward Genotype Editing in GA for several machine learning problems (Huang and Rocha, 2003, 2004). We show that the incorporation of genotype editing provides a means for artificial agents with genetic descriptions to gain greater phenotypic plasticity. Artificial agents use genotype edition to their advantage by linking it to environmental context. The ability to link changes in the environment with editing parameters gives organisms an adaptive advantage as genotype expression can become contextually regulated. The study of this RNA editing model in changing environments has shed some light into the evolutionary implications of RNA editing. We expect that our methodology will both facilitate determining the evolutionary role of RNA editing in biology, and advance the current state of research in Evolutionary Computation and Artificial Life.

1. RNA Editing

Evidence for the important role of non-protein coding RNA (ncRNA) in complex organisms (higher eukaryotes) has accumulated in recent years. "ncRNA dominates the genomic output of the higher organisms and has been shown to control chromosome architecture, mRNA turnover and the developmental timing of protein expression, and may also regulate transcription and alternative splicing." (Mattick, 2003, p 930).

RNA Editing (Benne, 1993; Bass, 2001), a process of post-transcriptional alteration of genetic information, can be performed by ncRNA structures (though it can also be performed by proteins). The term initially referred to the insertion or deletion of particular bases (e.g. uridine), or some sort of base conversion.

The most famous RNA editing system is that of the African Trypanosomes (Benne, 1993; Stuart, 1993). Its genetic material was found to possess strange sequence features such as genes without translational initiation and termination codons, frame shifted genes, etc. Furthermore, observation of mRNA's showed that many of them were significantly different from the genetic material from which they had been transcribed. These facts suggested that mRNA's were edited post-transcriptionally. It was later recognized that this editing was performed by guide RNA's (gRNA's) coded mostly by what was previously thought of as non-functional genetic material (Sturn and Simpson, 1990). In this particular genetic system, gRNA's operate by inserting, and sometimes deleting, uridines. To appreciate the effect of this edition let us consider Figure 1. The first example (Benne, 1993, p. 14) shows a massive uridine insertion (lowercase u's); the amino acid sequence that would be obtained prior to any edition is shown on top of the base sequence, and the amino acid sequence obtained after edition is shown in the gray box. The second example shows how, potentially, the insertion of a single uridine can change dramatically the amino acid sequence obtained; in this case, a termination codon is introduced.

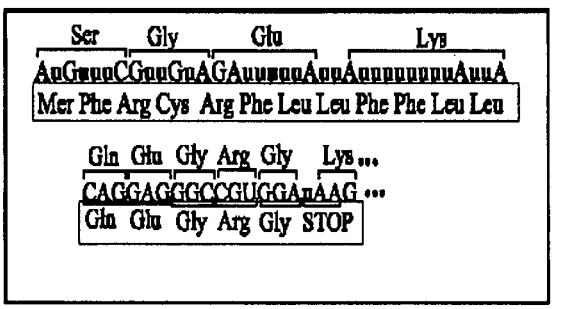

Figure 1. U-insertion in Trypanosomes' RNA

The importance of RNA Editing is thus unquestionable, since it has the power to dramatically alter gene expression: "cells with different mixes of (editing mechanisms) may edit a transcript from the same gene differently, thereby making different proteins from the same opened gene." (Pollack, 1994, P. 78). It is important to retain that a mRNA molecule can be more or less edited according to the concentrations of the editing operators it encounters. Thus, several different proteins coded by the same gene may coexist in an organism or even a cell, if all (or some) of the mRNA's obtained from the same gene, but edited differently, are meaningful to the translation mechanism.

If the concentrations of editing operators can be linked to environmental contexts, the concentrations of different proteins obtained may be selected accordingly, and thus evolve a system which is able to respond to environmental changes without changes in the major part of its genetic information -- one gene, different contexts, different proteins. This type of phenotypic plasticity may be precisely what the Trypanosome parasites have achieved: control over gene expression during different parts of their complex life cycles.

> "In mammalian genomes, gene duplication followed by separate evolution of the two copies would be a more obvious way of producing closely related proteins in regulatable amounts. RNA editing, however, does provide the opportunity to introduce highly specific, local changes into only some of the molecules. [...] It could be reasoned that somehow this would be more difficult to achieve via gene duplication, since independently accumulating mutations would make it harder to keep the remainder of the two sequences identical" (Benne, 1993, p. 22)

Thus, RNA editing may be more than just a system responsible for the introduction of environmentally regulated gene expression, but also a system that may allow the evolution of different proteins constrained by the same genetic string. In other words, even though one gene may produce different mRNA's (and thus proteins), the latter are not allowed heritable variation. What is inheritable, and subjected to variation, is the original non-edited gene, which is ultimately selected and transmitted to the offspring of the organism (Rocha, 1995; 1997).

The role of RNA editing in the development of more complex organisms has been shown to be important. Lomeli et al. (1994) discovered that the extent of RNA editing affecting a type of receptor channels responsible for the mediation of excitatory postsynaptic currents in the central nervous system, increases in rat brain development. As a consequence, the kinetic aspects of these channels differ according to the time of their creation in the brain's developmental process. Another example is that the development of rats without a gene (ADAR1) known to be involved in RNA editing, terminates midterm (Wang et al., 2000). This showed that RNA Editing is more prevalent and important than previously thought. RNA editing processes have also been identified in mammalian brains (Simpson and Emerson, 1996), including human brains (Mittaz et al., 1997). More recently, Hoopengardner et al. (2003) found that RNA editing plays a central role in nervous system function. Indeed, many edited sites recode conserved and functionally important amino acids, some of which may play a role in nervous system disorders such as epilepsy and Parkinson Disease.

2. Introducing Editing in Genetic Algorithms

Genetic Algorithms (GA) (Holland, 1975) have been used as computational models of natural evolutionary systems and as adaptive algorithms for solving optimization problems. GA operate on an evolving population of artificial organisms, or agents. Each agent is comprised of a genotype (encoding a solution to some problem) and a phenotype (the solution itself). Evolution occurs by iterated stochastic variation of genotypes, and selection of the best phenotypes in an environment according to how well the respective solution solves a problem (or fitness function).

Table 1 depicts the process of a simple genetic algorithm.

Table 1. Mechanism of a simple GA

1. Randomly generate an initial population of *l* *n*-bit agents, each defined by a genotype string (chromosome) of symbols from a small alphabet.
2. Evaluate each agent's (phenotype) fitness.
3. Repeat until *l* offspring agents have been created.
   a. select a pair of parent agents for mating;
   b. apply crossover operator to genotype string;
   c. apply mutation operator to genotype string.
4. Replace the current population with the new population.
5. Go to Step 2 until terminating condition.

The essence of GA lies on the separation of the description of a solution (the Genotype) from the solution itself (the Phenotype): variation is applied solely to the descriptions, while the respective solutions are evaluated, and the whole selected according to this evaluation. Nonetheless, one important difference between evolutionary computation and biological organisms, lies precisely on the relation between descriptions and solutions, between Genotype and Phenotype. In GA, typically, the relation between the two is linear and direct: one description, one solution. While in biological organisms there exists a multitude of processes, taking place between the transcription of a gene and its expression and subsequent development into a phenotype, responsible for the establishment of an uncertain, contextually regulated relation, between Genotype and Phenotype.

In other words, the same genotype will not always produce the same phenotype; rather, many phenotypes can be produced by one genotype depending on changes in the environment. For instance, in biological genetic systems with RNA editing, before a gene is translated into the space of proteins it may be altered through interactions with other types of molecules, namely RNA editors such as gRNA's.

If the effects of changing environmental context affecting gene expression within an individual can be harnessed and used to its selective advantage in a changing

environment, then we can say that such an individual has achieved a degree of control over its own genetic expression.

In analogy with the process of RNA Editing, Rocha (1995; 1997) proposed an expanded GA with stochastic edition of genotypes (chromosomes), prior to translation into phenotypes. Here we present novel experiments to show how this GA with Genotype Editing can be successfully used to model the environmentally-regulated control of gene expression achieved by RNA Editing in real organisms.

Genotype Editing (Rocha, 1995; Huang and Rocha, 2003, 2004) is implemented by a set of editors with different editing functions, such as insertion or deletion of symbols in the original chromosomes. Before chromosomes can be translated into the space of solutions, they must "pass" through successive layers of editors, present in different concentrations. In each generation, each chromosome has a certain probability (given by the concentrations) of encountering an editor in its layer. If an editor matches some subsequence of the chromosome when they encounter each other, the editor's function is applied and the chromosome is edited. The detailed implementation of the simplest GA with Edition (GAE) is described in the following:

The GAE model consists of a family of r m-bit strings, denoted as $(E_1, E_2, \ldots, E_r)$, which is used as the set of editors for the chromosomes of the agents in a GA population. The length of the editor strings is assumed much smaller than that of the chromosomes: $m << n$, usually an order of magnitude. An editor E_j is said to match a substring, of size m, of a chromosome, S, at position k if $e_i = s_{k+i}$, i=1,2, ..., m, $1 = k = n$-m, where e_i and s_i denote the i-th bit value of E_j and S, respectively. For each editor E_j, there exists an associated editing function F_j that specifies how a particular editor edits the chromosomes: when the editor matches a portion of a chromosome, a number of bits are inserted into or deleted from the chromosome.

For instance, if the editing function of editor E_j is to add one randomly generated allele at s_{k+m+1} when E_j matches S at position k, then all alleles of S from position $k+m+1$ to $n-1$ are shifted one position to the right (the allele at position n is removed). Analogously, if the editing function of editor E_j is to delete an allele, this editor will instead delete the allele at s_{k+m+1} when E_j matches S at position k. All the alleles after position $k+m+1$ are shifted in the inverse direction (one randomly generated allele is then assigned at position n).

Finally, let the concentration of the editor family be defined by $(\nu_1, \nu_2, \ldots, \nu_r)$. This means that the concentration of editor E_j is denoted by ν_j, and the probability that S encounters E_j is thus given by ν_j. With these settings, the algorithm for the GA with genotype editing is essentially the same as the regular GA, except that step 2 in Table 1 is now more complicated and redefined as:

> "For each individual in the GA population, apply each editor E_j with probability ν_j (i.e., concentration). If E_j matches the individual's chromosome S, then edit S with editing function F_j and evaluate the resulting individual's fitness."

It is important to notice that the "post-transcriptional" edition of genotypes is not a process akin to mutation, because editions are not inheritable. Just like in biological systems, it is the unedited genotype that is reproduced.

It is also important to retain that just like an mRNA molecule may be edited in different degrees according to the concentrations of editing operators it encounters, in the GAE the same chromosome may be edited differently because the editor concentration is a stochastic parameter that specifies the probability of a given editor encountering a chromosome. Thus, if a chromosome is repeated in the population, it may actually produce different solutions (or phenotypes). This mirrors what happens with RNA editing in biological organisms where, at the same time, several different proteins coded by the same gene may coexist.

In (Huang and Rocha, 2003, 2004), we have conducted a systematic study of the GAE in static environments to study if there are any evolutionary advantages of genotype editing, even without control of environmental changes. We demonstrated how the genotype editing can improve the GA's search performance by suppressing the effects of hitchhiking (Forrest and Mitchell, 1993). We have also showed that editing frequency plays a critical role in the evolutionary advantage provided by the editors -- only a moderate degree of editing processes facilitates the exploration of the search space. Therefore, one needs to choose proper editor parameters to avoid over or under-editions in order to develop more robust GAs. Here, we extend our study of the GAE to dynamic problems by linking concentrations of editors to environmental states (or contexts) – thus allowing editor concentrations to serve as a control switch for environmental changes.

3. Evolution in Dynamic Environments

How rapid is evolutionary change, and what determines the rates, patterns, and causes of change, or lack thereof? Answers to these questions can tell us much about the evolutionary process. The study of evolutionary rate in the context of GA usually involves defining performance measures that embody the idea of rate of adaptation, so that its change over time can be monitored for investigation.

In this paper, two evolutionary measures, the maximum fitness and the population fitness at each generation, are employed.[1] To understand how Genotype Editing works in

[1] The maximum fitness is the fitness of the best individual in the current population; the population fitness here is

the GAE model, we employ a testbed, the small Royal Road *S1* (Huang and Rocha, 2003) due to its simplicity for tracing the evolutionary advancement.

Table 2. Small royal road function *S*1

s_1 = 11111**********...**********; c_1 = 10
s_2 = *****11111**********...**********; c_2 = 10
s_3 = **********11111**********...**********; c_3 = 10
s_4 = **********11111**********...**********; c_4 = 10
s_5 = **********11111**********...**********; c_5 = 10
s_6 = **********11111**********...**********; c_6 = 10
s_7 = **********11111*****...**********; c_7 = 10
s_8 = **********11111; c_8 = 10

Table 2 illustrates the schematic of the small Royal Road function *S*1. This function involves a set of schemata $S = (s_1,...,s_8)$ and the fitness of a bit string (chromosome) x is defined as

$$F(x) = \sum_{s_i \in S} c_i \sigma_{s_i}(x),$$

where each c_i is a value assigned to the schema s_i as defined in the table; $\sigma_{s_i}(x)$ is defined as 1 if x is an instance of s_i and 0 otherwise. In this function, the fitness of the global optimum string (40 1's) is 10*8 = 80.

As a step towards the study of linking editors' concentrations with environmental contexts, we introduce another testbed (fitness landscape) in which each schema is comprised of all 0's and the other parameters remain the same as used in *S*1. The fitness landscapes consisting of schemata of all 1's and all 0's are called **L1** and **L0,** respectively. These two testbeds are maximally different in the configurations of their fitness landscapes. By oscillating these two landscapes, we are able to investigate the maximal effects of genotype editors in GAE.

Table 3. Parameters of the five editors

	Editor 1	Editor 2	Editor 3	Editor 4	Editor 5
Length	4	4	4	2	4
Alleles	1110	0011	0101	00	0111
Concentration	0.0635	0.0476	0.7302	0.2857	0.3175
Editing Function	Delete 4 bits	Add 3 bits	Delete 1 bit	Delete 3 bits	Delete 2 bits

The GAE experiments conducted in this section are based on a binary tournament selection, one-point crossover and mutation rates of 0.7 and 0.005, respectively; population size is 40 for each of 50 GAE runs. A family of 5 editors, C1, is randomly generated, with editor length selected in the range of 2 to 4 bits (see (Huang and Rocha, 2003, 2004) for a set of guidelines for parameter choices of the editors). Table 3 shows the corresponding parameters generated for each editor in family C1: length, alleles, concentration and editing function. For example, editor 3 is a bit-string of length 4 (0101); its concentration, or the probability that a chromosome will encounter this editor is 0.7302; its editing function is to delete 1 bit, meaning that this editor deletes 1 chromosome allele at the position following the chromosome substring that matches the editor's string.

Figure 2.a and 2.b display the averaged maximum fitness and averaged population fitness for the GAs and the GAEs on static environments L0 and L1, respectively.[2] In the figure, L0 (GA) and L1 (GA) denote the results obtained for the traditional GA on landscapes L0 and L1, respectively. L1C1 (GAE) denotes a GAE with the family of editors C1 shown in Table 3, applied to the L1 landscape. L0C1 (GAE) denotes a GAE with the same family of editors C1 applied to the L0 landscape.

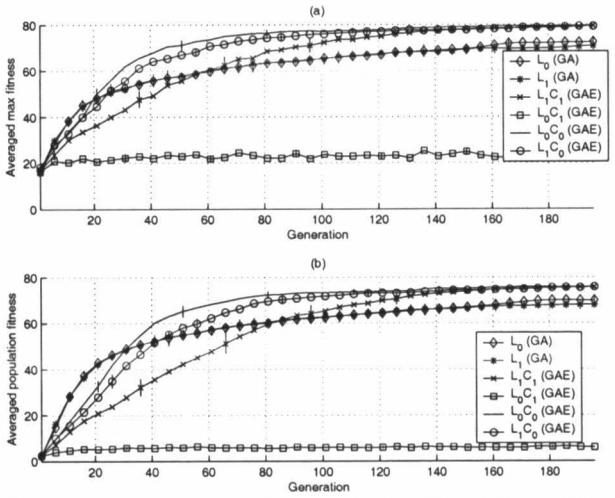

Figure 2. Evolutionary measures on static landscapes

One can see that the family of editors C1 facilitates the population's adaptation on L1 with respect to the maximum fitness and population fitness, in comparison with the traditional GA without edition on the same landscape. However, C1 is by no means beneficial for the GAE on landscape L0.

To enhance the performance of the GAE population on L0, we produced another editor family, C0, whose only difference from C1 is a new set of editor concentrations, {0.31, 0.062, 0.989, 0.002, 0.05}, with all other editor parameters remaining the same as in Table 3. The results in Figure 2 show that the GAE with C0 now performs much better on L0 than with C1. We also notice that the L1C1 and L0C0 GAE clearly outperform the GA without edition on L1 and L0 respectively.

defined as the value obtained by averaging the fitness of all the individuals in the current population.

[2] The value of the averaged maximum fitness measure is calculated by averaging the fitness of the best individuals at each generation for all 50 runs, where the vertical bars overlaying the measure curves represent the 95-percent confidence intervals. This applies to all the results obtained for the measures employed in this paper.

Consider now a dynamic environment which oscillates periodically between the landscapes L1 and L0. This oscillation models an environment with recurring dramatic changes in conditions. We know that some biological organisms, namely parasites that go through dramatic environmental changes, use the edition of mRNA molecules to their advantage, by linking the process of edition to environmental context. The ability to link changes in the environment with internal parameters such as concentrations of editing agents, is one of the mechanisms that can be used to (contextually) regulate gene expression (Mattick, 2003) with potential adaptive advantages (Rocha, 1995).

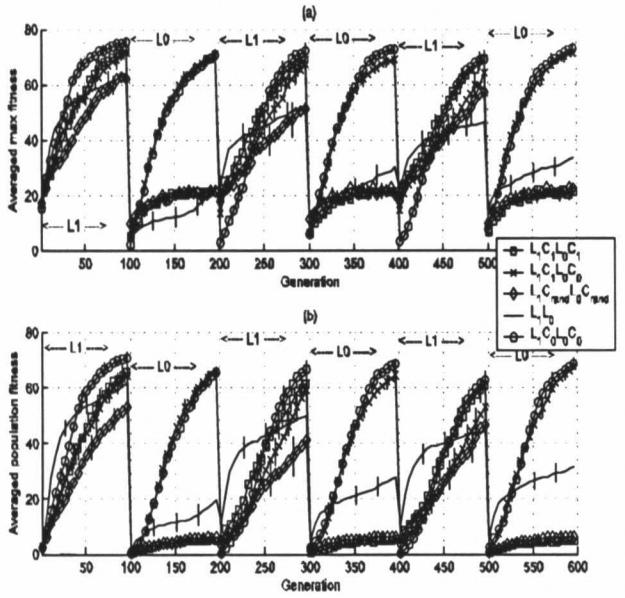

Figure 3. Evolutionary measures on dynamic landscapes

Figure 3 depicts our modeling of this process with the oscillation of landscapes L1 and L0, at every 100 generations. Several scenarios are tested:

- **L1L0**. Landscapes oscillate without genotype edition. The population evolves solely according to the traditional GA.
- **L1C1L0C1**. Landscapes oscillate with genotype edition, but edition is always implemented with family C1.
- **L1C0L0C0**. Same as above but with family C0.
- **L1C1L0C0**. Landscapes oscillate with edition, but the family of editors changes with the environment: family C1 operates when landscape L1 is in place, and C0 operates with L0.
- **L1C_{rand}L0C_{rand}**. Landscapes oscillate with genotype edition, but edition is always implemented with family C_{rand}. This is essentially the same as L1C1L0C1, except that the concentrations of editors are randomly generated at the start of each GAE run, and all other parameters are kept the same as in table 3.

The results for L1C1L0C0 show that the linking between the editor concentrations and environmental contexts (i.e., the linking between L1 and C1, and between L0 and C0) indeed provides adaptive advantages on the oscillating landscapes.

We also notice that C0 always advances the adaptation of the GAE's population even when the two landscapes oscillate (i.e., the results for L1C0L0C0). This means that family C0 is good at editing chromosomes in both landscapes. The results obtained for L1C1L0C0 and L1C0L0C0 show that genotype editing can lead to advantageous phenotypic plasticity in two ways to cope with dynamic environments: (1) by linking editors' concentrations with the environment (the case of L1C1L0C0), or (2) by employing editors which can produce chromosomes encoding good solutions in both landscapes (L1C0L0C0).

We do notice, however, that strategy 1 provides a quicker response immediately after the environment changes from L0 to L1. In figure 3, we can see that when this change occurs, the maximum fitness of L1C0L0C0 suffers a larger setback than that of L1C1L0C0; that is, the population of the first needs to completely re-adapt to the new environment, whereas the population of the second contains some elements already with moderate fitness when the landscape changes.

A microscopic inspection shows that in the case of L1C1L0C0, at generation 199, the chromosome of one individual of fitness 60 is defined by substring {0,1,0,1,0} at the position of schema $S7$. When the landscape oscillates from L0 to L1 at generation 200, this individual undergoes some edition which results in these alleles being altered to {1,1,1,1,1}. This individual thus acquires a fitness amount of 10 from building block $S7$. This situation is relatively typical in L1C1L0C0; yet in L1C0L0C0, since more individuals converge to all 0's at generation 199, it is therefore more difficult for the population individuals to be able to acquire corresponding building blocks at generation 200 simply by genotype edition. All this means that under strategy 2, the GAE evolves chromosomes which produce fair solutions in both landscapes, but which are edited differently accordingly. Therefore, the same chromosomes may exist in both landscapes, whereas in the case of strategy 1, C0 seems to facilitate the evolution of new chromosomes every time the landscape changes.

4. Conclusion and Future Work

This paper presents our computational methodology using Genetic Algorithms with Genotype Editing for investigating the role of RNA editing in dynamic environments. Based on several genetic editing characteristics that are gleaned from the RNA editing system, we show that the incorporation of editing

mechanisms indeed provides a means for artificial agents with genetic descriptions to gain greater phenotypic plasticity. By linking changes in the environment with internal parameters such as concentrations of editors, the artificial agents can use genotype edition to their advantage, as gene expression can become contextually regulated, such ability thus gives organisms an adaptive advantage. In a nutshell, the results obtained have provided the following insights:

There are two strategies for the artificial organisms with genotype edition to produce phenotypic plasticity to cope with environmental changes: (1) by using different families of editors for different environmental demands, or (2) by employing a single family of editors that allows the evolutionary process to cope well with a changing environment.

We have thus far studied the linking of editor families with different concentrations to external contextual changes. In future work, we intend to allow the family of editors and the agents' genotype to co-evolve, so that the artificial agents can discover proper editor concentrations to adapt to changing environments. Since there are several internal editor parameters involved in an editing system, such as the size of the editor family, editor length and editor functions, in addition to the investigation of editor concentrations, our future work is also going to study the effects of linking other parameters with external environments. Since the length of oscillation period is expected to be another critical parameter that will affect how well the GAE's population adapts to changing environments, we will also study the effects of oscillation periods. With a systematic study on these editor parameters, our hope is to gain a deeper understanding of the role of RNA Editing in nature and also to design robust evolutionary computation algorithms for complex, real-world tasks (as we have done in Huang and Rocha 2003, 2003).

References

Bass, B. L. (Ed.) (2001). RNA Editing. Frontiers in Molecular Biology Series. Oxford University Press.

Benne, R. (Ed.) (1993). RNA Editing: The Alteration of Protein Coding Sequences of RNA. Ellis Horwood.

Forrest, S. and Mitchell, M. (1993). "Relative Building Block Fitness and the Building Block Hypothesis." Foundations of Genetic Algorithms 2: 109-126.

Holland, J. H. (1975). Adaptation in Natural and Artificial Systems. University of Michigan Press.

Hoopengardner, B., Bhalla, T., Staber, C., and Reenan, R. (2003). "Nervous System Targets of RNA Editing Identified by Comparative Genomics." Science 301: 832-836.

Huang, C.-F. and Rocha, L. M. (2003). "Exploration of RNA Editing and Design of Robust Genetic Algorithms." Proc. 2003 Congress on Evolutionary Computation, Canberra, Australia, December 2003. R. Sarker et al (Eds). IEEE Press, pp. 2799-2806.

Huang, C.-F. and Rocha, L. M. (2004). "A Systematic Study of Genetic Algorithms with Genotype Editing." *GECCO 2004*. Submitted.

Langton, C.G. (1989). "Artificial life." In: *Artificial Life*. C. Langton (ed.). Addison-Wesley.

Lomeli, H. et al. (1994). "Control of Kinetic Properties of AMPA Receptor Channels by RNA Editing." Science, 266: 1709-1713.

Mattick, J. S. (2003). Challenging the Dogma: the Hidden Layer of Non-protein-coding RNAs in Complex Organisms." BioEssays. 25: 930-939.

Mittaz, L., Antonarakis, S. E., Higuichi, M. and Scott, H. S. (1997). "Localization of a Novel Human RNA-editing Deaminase (hRED2 or ADARB2) to Chromosome 10p15." Human Genetics, 100: 398-400.

Pollack, R. (1994). Signs of Life: The Language and Meanings of DNA. Houghton Mifflin.

Rocha, Luis M. (1995). "Contextual Genetic Algorithms: Evolving Developmental Rules." Advances in Artificial Life. Moran, J., Moreno, A., Merelo, J. J., and Chacon, P. (Eds.). Springer Verlag, pp. 368-382.

Rocha, Luis M. (1997). Evidence Sets and Contextual Genetic Algorithms: Exploring Uncertainty, Context and Embodiment in Cognitive and biological Systems. PhD. Dissertation. State University of New York at Binghamton. Science.

Simpson, L. and Emerson, R. B. (1996). "RNA Editing." Annual Review of Neuroscience, 19: 27-52.

Stuart, K. (1993). "RNA Editing in Mitochondria of African Trypanosomes." In: RNA Editing : The Alteration of Protein Coding Sequences of RNA. Benne, R. (Ed.). Ellis Horwood Publishers. pp. 26-52.

Sturn, N. R. and Simpson, L. (1990). "Kinetoplast DNA Minicircles Encode Guide RNA's for Editing of Cytochrome Oxidase Subunit III mRNA." Cell, 61: 879-884.

Wang, Q., Khillan, J., Gadue, P., and Nishikura, K. (2000). "Requirement of the RNA Editing Deaminase ADAR1 Gene for Embryonic Erythropoiesis." Science, 290 (5497): 1765-1768.